

## The Modern Data Warehouse A New Approach For A New Era

Enter the fast-paced world of SAP HANA 2.0 with this introductory guide. Begin with an exploration of the technological backbone of SAP HANA as a database and platform. Then, step into key SAP HANA user roles and discover core capabilities for administration, application development, advanced analytics, security, data integration, and more. No matter how SAP HANA 2.0 fits into your business, this book is your starting point. In this book, you'll learn about:

- Technology Discover what makes an in-memory database platform. Learn about SAP HANA's journey from version 1.0 to 2.0, take a tour of your technology options, and walk through deployment scenarios and implementation requirements.
- Tools Unpack your SAP HANA toolkit. See essential tools in action, from SAP HANA cockpit and SAP HANA studio, to the SAP HANA Predictive Analytics Library and SAP HANA smart data integration.
- Key Roles Understand how to use SAP HANA as a developer, administrator, data scientist, data center architect, and more. Explore key tasks like backend programming with SQLScript, security setup with roles and authorizations, data integration with the SAP HANA Data Management Suite, and more. Highlights include: 1) Architecture 2) Administration 3) Application development 4) Analytics 5) Security 6) Data integration 7) Data architecture 8) Data center

The data lakehouse is the next generation of the data warehouse and data lake, designed to meet today's complex and ever-changing analytics, machine learning, and data science requirements. Learn about the features and architecture of the data lakehouse, along with its powerful analytical infrastructure. Appreciate how the universal common connector blends structured, textual, analog, and IoT data. Maintain the lakehouse for future generations through Data Lakehouse Housekeeping and Data Future-proofing. Know how to incorporate the lakehouse into an existing data governance strategy. Incorporate data catalogs, data lineage tools, and open source software into your architecture to ensure your data scientists, analysts, and end users live happily ever after.

Leverage the power of Microsoft Azure Data Factory v2 to build hybrid data solutions Key Features Combine the power of Azure Data Factory v2 and SQL Server Integration Services Design and enhance performance and scalability of a modern ETL hybrid solution Interact with the loaded data in data warehouse and data lake using Power BI Book Description ETL is one of the essential techniques in data processing. Given data is everywhere, ETL will always be the vital process to handle data from different sources. Hands-On Data Warehousing with Azure Data Factory starts with the basic concepts of data warehousing and ETL process. You will learn how Azure Data Factory and SSIS can be used to understand the key components of an ETL solution. You will go through different services offered by Azure that can be used by ADF and SSIS, such as Azure Data Lake Analytics, Machine Learning and Databrick's Spark with the help of practical examples. You will explore how to design and implement ETL hybrid solutions using different integration services with a step-by-step approach. Once you get to grips with all this, you will use Power BI to interact with data coming from different sources in order to reveal valuable insights. By the end of this book, you will not only learn how to build your own ETL solutions but also address the key challenges that are faced while building them. What

## Get Free The Modern Data Warehouse A New Approach For A New Era

you will learn Understand the key components of an ETL solution using Azure Data Factory and Integration Services Design the architecture of a modern ETL hybrid solution Implement ETL solutions for both on-premises and Azure data Improve the performance and scalability of your ETL solution Gain thorough knowledge of new capabilities and features added to Azure Data Factory and Integration Services Who this book is for This book is for you if you are a software professional who develops and implements ETL solutions using Microsoft SQL Server or Azure cloud. It will be an added advantage if you are a software engineer, DW/ETL architect, or ETL developer, and know how to create a new ETL implementation or enhance an existing one with ADF or SSIS.

Foreword by Mark Stephen LaRow, Vice President of Products, MicroStrategy "A unique and authoritative book that blends recent research developments with industry-level practices for researchers, students, and industry practitioners." Il-Yeol Song, Professor, College of Information Science and Technology, Drexel University

Get started with Azure Synapse Analytics, Microsoft's modern data analytics platform. This book covers core components such as Synapse SQL, Synapse Spark, Synapse Pipelines, and many more, along with their architecture and implementation. The book begins with an introduction to core data and analytics concepts followed by an understanding of traditional/legacy data warehouse, modern data warehouse, and the most modern data lakehouse. You will go through the introduction and background of Azure Synapse Analytics along with its main features and key service capabilities. Core architecture is discussed, along with Synapse SQL. You will learn its main features and how to create a dedicated Synapse SQL pool and analyze your big data using Serverless Synapse SQL Pool. You also will learn Synapse Spark and Synapse Pipelines, with examples. And you will learn Synapse Workspace and Synapse Studio followed by Synapse Link and its features. You will go through use cases in Azure Synapse and understand the reference architecture for Synapse Analytics. After reading this book, you will be able to work with Azure Synapse Analytics and understand its architecture, main components, features, and capabilities. What You Will Learn Understand core data and analytics concepts and data lakehouse concepts Be familiar with overall Azure Synapse architecture and its main components Be familiar with Synapse SQL and Synapse Spark architecture components Work with integrated Apache Spark (aka Synapse Spark) and Synapse SQL engines Understand Synapse Workspace, Synapse Studio, and Synapse Pipeline Study reference architecture and use cases Who This Book Is For Azure data analysts, data engineers, data scientists, and solutions architects

This book contains practical steps business users can take to implement data management in a number of ways, including data governance, data architecture, master data management, business intelligence, and others. It defines data strategy, and covers chapters that illustrate how to align a data strategy with the business strategy, a discussion on valuing data as an asset, the evolution of data management, and who should oversee a data strategy. This provides the user with a good understanding of what a data strategy is and its limits. Critical to a data strategy is the incorporation of one or more data management domains. Chapters on key data management domains—data governance, data architecture, master data management and analytics, offer the user a

## Get Free The Modern Data Warehouse A New Approach For A New Era

practical approach to data management execution within a data strategy. The intent is to enable the user to identify how execution on one or more data management domains can help solve business issues. This book is intended for business users who work with data, who need to manage one or more aspects of the organization's data, and who want to foster an integrated approach for how enterprise data is managed. This book is also an excellent reference for students studying computer science and business management or simply for someone who has been tasked with starting or improving existing data management.

Agile Data Warehouse Design is a step-by-step guide for capturing data warehousing/business intelligence (DW/BI) requirements and turning them into high performance dimensional models in the most direct way: by modelstorming (data modeling ] brainstorming) with BI stakeholders. This book describes BEAM, an agile approach to dimensional modeling, for improving communication between data warehouse designers, BI stakeholders and the whole DW/BI development team. BEAM provides tools and techniques that will encourage DW/BI designers and developers to move away from their keyboards and entity relationship based tools and model interactively with their colleagues. The result is everyone thinks dimensionally from the outset! Developers understand how to efficiently implement dimensional modeling solutions. Business stakeholders feel ownership of the data warehouse they have created, and can already imagine how they will use it to answer their business questions. Within this book, you will learn: Agile dimensional modeling using Business Event Analysis & Modeling (BEAM ) Modelstorming: data modeling that is quicker, more inclusive, more productive, and frankly more fun! Telling dimensional data stories using the 7Ws (who, what, when, where, how many, why and how) Modeling by example not abstraction; using data story themes, not crow's feet, to describe detail Storyboarding the data warehouse to discover conformed dimensions and plan iterative development Visual modeling: sketching timelines, charts and grids to model complex process measurement - simply Agile design documentation: enhancing star schemas with BEAM dimensional shorthand notation Solving difficult DW/BI performance and usability problems with proven dimensional design patterns LawrenceCorr is a data warehouse designer and educator. As Principal of DecisionOne Consulting, he helps clients to review and simplify their data warehouse designs, and advises vendors on visual data modeling techniques. He regularly teaches agile dimensional modeling courses worldwide and has taught dimensional DW/BI skills to thousands of students. Jim Stagnitto is a data warehouse and master data management architect specializing in the healthcare, financial services, and information service industries. He is the founder of the data warehousing and data mining consulting firm Llumino.

With constantly expanding options such as Azure Data Lake Server (ADLS) and Azure SQL Data Warehouse (ADW), how can developers learn the process and components required to successfully move this data? Quick Start Guide to Azure Data Factory, Azure Data Lake Server, and Azure Data Warehouse teaches you the basics of moving data between Azure SQL solutions using Azure Data Factory. Discover how to build and deploy each of the components needed to integrate data in the cloud with local SQL databases. Mark Beckner's step by step instructions on how to build each component, how to test processes and debug, and how to track and audit the movement of data, will help you to build your own solutions instantly and efficiently. This book includes

## Get Free The Modern Data Warehouse A New Approach For A New Era

information on configuration, development, and administration of a fully functional solution and outlines all of the components required for moving data from a local SQL instance through to a fully functional data warehouse with facts and dimensions. Leverage the Azure analytics platform's key analytics services to deliver unmatched intelligence for your data Key Features Learn to ingest, prepare, manage, and serve data for immediate business requirements Bring enterprise data warehousing and big data analytics together to gain insights from your data Develop end-to-end analytics solutions using Azure Synapse Book Description Azure Synapse Analytics, which Microsoft describes as the next evolution of Azure SQL Data Warehouse, is a limitless analytics service that brings enterprise data warehousing and big data analytics together. With this book, you'll learn how to discover insights from your data effectively using this platform. The book starts with an overview of Azure Synapse Analytics, its architecture, and how it can be used to improve business intelligence and machine learning capabilities. Next, you'll go on to choose and set up the correct environment for your business problem. You'll also learn a variety of ways to ingest data from various sources and orchestrate the data using transformation techniques offered by Azure Synapse. Later, you'll explore how to handle both relational and non-relational data using the SQL language. As you progress, you'll perform real-time streaming and execute data analysis operations on your data using various languages, before going on to apply ML techniques to derive accurate and granular insights from data. Finally, you'll discover how to protect sensitive data in real time by using security and privacy features. By the end of this Azure book, you'll be able to build end-to-end analytics solutions while focusing on data prep, data management, data warehousing, and AI tasks. What you will learn Explore the necessary considerations for data ingestion and orchestration while building analytical pipelines Understand pipelines and activities in Synapse pipelines and use them to construct end-to-end data-driven workflows Query data using various coding languages on Azure Synapse Focus on Synapse SQL and Synapse Spark Manage and monitor resource utilization and query activity in Azure Synapse Connect Power BI workspaces with Azure Synapse and create or modify reports directly from Synapse Studio Create and manage IP firewall rules in Azure Synapse Who this book is for This book is for data engineers, IT professionals, business analysts, data scientists, and database administrators who are looking to get up and running with the Azure Synapse Analytics platform. Basic knowledge of data warehousing will be beneficial to help you understand the concepts covered in this book more effectively.

Explore the modern market of data analytics platforms and the benefits of using Snowflake computing, the data warehouse built for the cloud. With the rise of cloud technologies, organizations prefer to deploy their analytics using cloud providers such as Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform. Cloud vendors are offering modern data platforms for building cloud analytics solutions to collect data and consolidate into single storage

## Get Free The Modern Data Warehouse A New Approach For A New Era

solutions that provide insights for business users. The core of any analytics framework is the data warehouse, and previously customers did not have many choices of platform to use. Snowflake was built specifically for the cloud and it is a true game changer for the analytics market. This book will help onboard you to Snowflake, present best practices to deploy, and use the Snowflake data warehouse. In addition, it covers modern analytics architecture and use cases. It provides use cases of integration with leading analytics software such as Matillion ETL, Tableau, and Databricks. Finally, it covers migration scenarios for on-premise legacy data warehouses. What You Will Learn Know the key functionalities of Snowflake Set up security and access with cluster Bulk load data into Snowflake using the COPY command Migrate from a legacy data warehouse to Snowflake integrate the Snowflake data platform with modern business intelligence (BI) and data integration tools Who This Book Is For Those working with data warehouse and business intelligence (BI) technologies, and existing and potential Snowflake users

Analyze vast amounts of data in record time using Apache Spark with Databricks in the Cloud. Learn the fundamentals, and more, of running analytics on large clusters in Azure and AWS, using Apache Spark with Databricks on top. Discover how to squeeze the most value out of your data at a mere fraction of what classical analytics solutions cost, while at the same time getting the results you need, incrementally faster. This book explains how the confluence of these pivotal technologies gives you enormous power, and cheaply, when it comes to huge datasets. You will begin by learning how cloud infrastructure makes it possible to scale your code to large amounts of processing units, without having to pay for the machinery in advance. From there you will learn how Apache Spark, an open source framework, can enable all those CPUs for data analytics use. Finally, you will see how services such as Databricks provide the power of Apache Spark, without you having to know anything about configuring hardware or software. By removing the need for expensive experts and hardware, your resources can instead be allocated to actually finding business value in the data. This book guides you through some advanced topics such as analytics in the cloud, data lakes, data ingestion, architecture, machine learning, and tools, including Apache Spark, Apache Hadoop, Apache Hive, Python, and SQL. Valuable exercises help reinforce what you have learned. What You Will Learn Discover the value of big data analytics that leverage the power of the cloud Get started with Databricks using SQL and Python in either Microsoft Azure or AWS Understand the underlying technology, and how the cloud and Apache Spark fit into the bigger picture See how these tools are used in the real world Run basic analytics, including machine learning, on billions of rows at a fraction of a cost or free Who This Book Is For Data engineers, data scientists, and cloud architects who want or need to run advanced analytics in the cloud. It is assumed that the reader has data experience, but perhaps minimal exposure to Apache Spark and Azure Databricks. The book is also recommended for people who want to get started in the analytics field, as it

## Get Free The Modern Data Warehouse A New Approach For A New Era

provides a strong foundation.

Leverage the power of Azure to get efficient data insights from your big data in real time

### Key Features

Explore the basics of cloud analytics using Azure Discover different ways to process and visualize your data easily Learn to use Azure Synapse Analytics (formerly known as Azure SQL Data Warehouse) to derive real-time customer insights

### Book Description

With data being generated at an exponential speed, organizations all over the world are migrating their infrastructure to the cloud. Application management becomes much easier when you use a cloud platform to build, manage, and deploy your services and applications. Cloud Analytics with Microsoft Azure covers all that you need to extract useful insights from your data. You'll explore the power of data with big data analytics, the Internet of Things (IoT), machine learning, artificial intelligence, and DataOps. You'll also delve into data analytics by studying use cases that focus on creating actionable insights from near-real-time data. As you advance, you'll learn to build an end-to-end analytics pipeline on the cloud with machine learning and deep learning concepts. By the end of this book, you'll have developed a solid understanding of data analytics with Azure and its practical implementation. What you will learn

- Explore the concepts of modern data warehouses and data pipelines
- Discover different design considerations while applying a cloud analytics solution
- Design an end-to-end analytics pipeline on the cloud
- Differentiate between structured, semi-structured, and unstructured data
- Choose a cloud-based service for your data analytics solutions
- Use Azure services to ingest, store and analyze data of any scale

Who this book is for If you're planning to adopt the cloud analytics model for your business, this book will help you understand the design and business considerations that you must keep in mind. Though not necessary, a basic understanding of data analytics concepts such as data streaming, data types, the machine learning life cycle, and Docker containers will help you get the most out of the book.

Data pipelines are the foundation for success in data analytics. Moving data from numerous diverse sources and transforming it to provide context is the difference between having data and actually gaining value from it. This pocket reference defines data pipelines and explains how they work in today's modern data stack. You'll learn common considerations and key decision points when implementing pipelines, such as batch versus streaming data ingestion and build versus buy. This book addresses the most common decisions made by data professionals and discusses foundational concepts that apply to open source frameworks, commercial products, and homegrown solutions. You'll learn:

- What a data pipeline is and how it works
- How data is moved and processed on modern data infrastructure, including cloud platforms
- Common tools and products used by data engineers to build pipelines
- How pipelines support analytics and reporting needs
- Considerations for pipeline maintenance, testing, and alerting

Data Warehousing in the Age of the Big Data will help you and your organization make the most of unstructured data with

## Get Free The Modern Data Warehouse A New Approach For A New Era

your existing data warehouse. As Big Data continues to revolutionize how we use data, it doesn't have to create more confusion. Expert author Krish Krishnan helps you make sense of how Big Data fits into the world of data warehousing in clear and concise detail. The book is presented in three distinct parts. Part 1 discusses Big Data, its technologies and use cases from early adopters. Part 2 addresses data warehousing, its shortcomings, and new architecture options, workloads, and integration techniques for Big Data and the data warehouse. Part 3 deals with data governance, data visualization, information life-cycle management, data scientists, and implementing a Big Data-ready data warehouse. Extensive appendixes include case studies from vendor implementations and a special segment on how we can build a healthcare information factory. Ultimately, this book will help you navigate through the complex layers of Big Data and data warehousing while providing you information on how to effectively think about using all these technologies and the architectures to design the next-generation data warehouse. Learn how to leverage Big Data by effectively integrating it into your data warehouse. Includes real-world examples and use cases that clearly demonstrate Hadoop, NoSQL, HBASE, Hive, and other Big Data technologies Understand how to optimize and tune your current data warehouse infrastructure and integrate newer infrastructure matching data processing workloads and requirements Data warehousing is one of the hottest business topics, and there's more to understanding data warehousing technologies than you might think. Find out the basics of data warehousing and how it facilitates data mining and business intelligence with Data Warehousing For Dummies, 2nd Edition. Data is probably your company's most important asset, so your data warehouse should serve your needs. The fully updated Second Edition of Data Warehousing For Dummies helps you understand, develop, implement, and use data warehouses, and offers a sneak peek into their future. You'll learn to: Analyze top-down and bottom-up data warehouse designs Understand the structure and technologies of data warehouses, operational data stores, and data marts Choose your project team and apply best development practices to your data warehousing projects Implement a data warehouse, step by step, and involve end-users in the process Review and upgrade existing data storage to make it serve your needs Comprehend OLAP, column-wise databases, hardware assisted databases, and middleware Use data mining intelligently and find what you need Make informed choices about consultants and data warehousing products Data Warehousing For Dummies, 2nd Edition also shows you how to involve users in the testing process and gain valuable feedback, what it takes to successfully manage a data warehouse project, and how to tell if your project is on track. You'll find it's the most useful source of data on the topic!

The data warehousing bible updated for the new millennium Updated and expanded to reflect the many technological advances occurring since the previous edition, this latest edition of the data warehousing "bible" provides a

## Get Free The Modern Data Warehouse A New Approach For A New Era

comprehensive introduction to building data marts, operational data stores, the Corporate Information Factory, exploration warehouses, and Web-enabled warehouses. Written by the father of the data warehouse concept, the book also reviews the unique requirements for supporting e-business and explores various ways in which the traditional data warehouse can be integrated with new technologies to provide enhanced customer service, sales, and support-both online and offline-including near-line data storage techniques.

Tips, techniques, and trends on how to use dashboard technology to optimize business performance Business performance management is a hot new management discipline that delivers tremendous value when supported by information technology. Through case studies and industry research, this book shows how leading companies are using performance dashboards to execute strategy, optimize business processes, and improve performance. Wayne W. Eckerson (Hingham, MA) is the Director of Research for The Data Warehousing Institute (TDWI), the leading association of business intelligence and data warehousing professionals worldwide that provide high-quality, in-depth education, training, and research. He is a columnist for SearchCIO.com, DM Review, Application Development Trends, the Business Intelligence Journal, and TDWI Case Studies & Solution.

The Data Vault was invented by Dan Linstedt at the U.S. Department of Defense, and the standard has been successfully applied to data warehousing projects at organizations of different sizes, from small to large-size corporations. Due to its simplified design, which is adapted from nature, the Data Vault 2.0 standard helps prevent typical data warehousing failures. "Building a Scalable Data Warehouse" covers everything one needs to know to create a scalable data warehouse end to end, including a presentation of the Data Vault modeling technique, which provides the foundations to create a technical data warehouse layer. The book discusses how to build the data warehouse incrementally using the agile Data Vault 2.0 methodology. In addition, readers will learn how to create the input layer (the stage layer) and the presentation layer (data mart) of the Data Vault 2.0 architecture including implementation best practices. Drawing upon years of practical experience and using numerous examples and an easy to understand framework, Dan Linstedt and Michael Olschimke discuss: How to load each layer using SQL Server Integration Services (SSIS), including automation of the Data Vault loading processes. Important data warehouse technologies and practices. Data Quality Services (DQS) and Master Data Services (MDS) in the context of the Data Vault architecture. Provides a complete introduction to data warehousing, applications, and the business context so readers can get-up and running fast Explains theoretical concepts and provides hands-on instruction on how to build and implement a data warehouse Demystifies data vault modeling with beginning, intermediate, and advanced techniques Discusses the advantages of the data vault approach over other techniques, also including the latest updates to Data Vault 2.0 and multiple improvements to Data Vault 1.0

The "father of data warehousing" incorporates the latest technologies into his blueprint for integrated decision support systems Today's corporate IT and data warehouse managers are required to make a small army of technologies work together to ensure fast and accurate information for business managers. Bill Inmon created the Corporate Information Factory to solve the needs of these managers. Since the First Edition, the design of the factory has grown and changed dramatically. This Second Edition,

## Get Free The Modern Data Warehouse A New Approach For A New Era

revised and expanded by 40% with five new chapters, incorporates these changes. This step-by-step guide will enable readers to connect their legacy systems with the data warehouse and deal with a host of new and changing technologies, including Web access mechanisms, e-commerce systems, ERP (Enterprise Resource Planning) systems. The book also looks closely at exploration and data mining servers for analyzing customer behavior and departmental data marts for finance, sales, and marketing.

With the help of well-structured and practical recipes, this book will teach you how to integrate data from the cloud and on-premise. You'll learn how to transform, clean, and consolidate data into a single data platform and get to grips with using ADF as the main ETL and orchestration tool for your data warehouse or data platform project.

There are more than one billion documents on the Web, with the count continually rising at a pace of over one million new documents per day. As information increases, the motivation and interest in data warehousing and mining research and practice remains high in organizational interest. The Encyclopedia of Data Warehousing and Mining, Second Edition, offers thorough exposure to the issues of importance in the rapidly changing field of data warehousing and mining. This essential reference source informs decision makers, problem solvers, and data mining specialists in business, academia, government, and other settings with over 300 entries on theories, methodologies, functionalities, and applications.

"This book provides insight into the latest findings concerning data warehousing, data mining, and their applications in everyday human activities"--Provided by publisher.

This book helps you to understand Snowflake's unique architecture and ecosystem that places it at the forefront of cloud data warehouses. The recipes present in this book will enable you to develop proficiency in managing data on Snowflake and learn Snowflake's novel features such as data sharing, cloning, and time travel.

Build a modern data warehouse on Microsoft's Azure Platform that is flexible, adaptable, and fast—fast to snap together, reconfigure, and fast at delivering results to drive good decision making in your business. Gone are the days when data warehousing projects were lumbering dinosaur-style projects that took forever, drained budgets, and produced business intelligence (BI) just in time to tell you what to do 10 years ago. This book will show you how to assemble a data warehouse solution like a jigsaw puzzle by connecting specific Azure technologies that address your own needs and bring value to your business. You will see how to implement Lambda architecture by combining batch ETL/ELT jobs using Azure Data Factory with streaming technologies such as EventHub and Azure Databricks. You will discover how to manage cleansing and transformation prior to serving your blended datasets up to queries arriving from Azure SQL Data Warehouse. And you will know how to build robust analytic solutions on Power BI and Azure Analysis Services to empower data-driven decision making that drives your business forward toward a pattern of success. This book teaches you how to employ the Azure platform in a strategy to dramatically improve implementation speed and flexibility of data warehousing systems. You will know how to make correct decisions in design, architecture, and infrastructure such as choosing which type of SQL engine (from at least three options) best

## Get Free The Modern Data Warehouse A New Approach For A New Era

meets the needs of your organization. You also will learn about ETL/ELT structure and the vast number of accelerators and patterns that can be used to aid implementation and ensure resilience. Data warehouse developers and architects will find this book a tremendous resource for moving their skills into the future through cloud-based implementations. What You Will Learn

- Choose the appropriate Azure SQL engine for implementing a given data warehouse
- Develop smart, reusable ETL/ELT processes that are resilient and easily maintained
- Automate mundane development tasks through tools such as PowerShell
- Ensure consistency of data by creating and enforcing data contracts
- Explore streaming and event-driven architectures for data ingestion
- Create advanced staging layers using Azure Data Lake Gen 2 to feed your data warehouse

Who This Book Is For Data warehouse or ETL/ELT developers who wish to implement a data warehouse project in the Azure cloud, and developers currently working in on-premise environments who want to move to the cloud, and for developers with Azure experience looking to tighten up their implementation and consolidate their knowledge

Data Warehouse projects fail. As an industry we have been battling with this phenomenon for decades. Though we have been getting better over the years, as an industry we still have a long way to go. Fortunately some people have found ways to beat the odds. By thinking out of the box, formulating new ideas, and creating new innovative approaches these people have each somehow unlocked the secrets of successful DW programs. Within this group there are those who have fire tested their theories with real life deployments. They have proven the merits of their ideas with perhaps the only test that really matters – actually deploying successful DW programs. Now this is a small group, a special club, and of the few members in this club, an even smaller group is willing and able to communicate and share their ideas. In this book, Integrated Data Hub™, Dario Mangano shares with us his ideas including the solution he developed and the proven architecture that he has successfully deployed in several organizations. This book is a guide to a successful data warehousing business intelligence (DWBI) program. One of the success factors is recognizing that there is no success to be found in a data warehouse project. To leverage enterprise data we have to recognize two things; first that the data warehouse is not a project with a beginning and end but rather an ongoing program, and second that this program is not technically driven but actually business and information driven. The Integrated Data Hub™ is an organizational information framework that takes into account the dynamic characteristics of the data warehouse function. Business alignment, data integration and historization are ongoing dynamic functions that need to respond to changes in the business, changes in the sources, and changes in downstream requirements. The organizational need for enterprise-wide data is an increasingly important component of enterprise management. With this increasing importance comes also an increasing need for traceability and auditability. Dario explains how the role of the business intelligence department is to manage the transformation of integrated and trusted data into understandable information with real business meaning. Dario's IDH

## Get Free The Modern Data Warehouse A New Approach For A New Era

leverages the concepts and principles of both Bill Inmon's data warehouse and corporate information factory, and of Ralph Kimball's dimensional modeling and data marts. Insofar as the goals of the IDH require an optimized modeling paradigm, Dario has specified the use of the data vault modeling approach. Dario starts by sketching the historical background and then introduces the IDH concept. Keep reading as Dario weaves the IDH components together into an effective strategy and a cohesive blueprint for enterprise information management. From a modeling perspective this book takes us from 3NF and Star Schemas to Data Vault and Dario's own Leaf Schemas. From an architectural perspective this book describes the component layers and their specific characteristics. This book is a great addition to the IP for the new age of data warehousing and information management. Data Warehouse programs using the IDH™ succeed. And perhaps that is the most compelling reason to read this book. Hans Hultgren (Author of the book: Modeling the Agile Data Warehouse with Data Vault)

Build a modern data warehouse on Microsoft's Azure Platform that is flexible, adaptable, and fast—fast to snap together, reconfigure, and fast at delivering results to drive good decision making in your business. Gone are the days when data warehousing projects were lumbering dinosaur-style projects that took forever, drained budgets, and produced business intelligence (BI) just in time to tell you what to do 10 years ago. This book will show you how to assemble a data warehouse solution like a jigsaw puzzle by connecting specific Azure technologies that address your own needs and bring value to your business. You will see how to implement a range of architectural patterns using batches, events, and streams for both data lake technology and SQL databases. You will discover how to manage metadata and automation to accelerate the development of your warehouse while establishing resilience at every level. And you will know how to feed downstream analytic solutions such as Power BI and Azure Analysis Services to empower data-driven decision making that drives your business forward toward a pattern of success. This book teaches you how to employ the Azure platform in a strategy to dramatically improve implementation speed and flexibility of data warehousing systems. You will know how to make correct decisions in design, architecture, and infrastructure such as choosing which type of SQL engine (from at least three options) best meets the needs of your organization. You also will learn about ETL/ELT structure and the vast number of accelerators and patterns that can be used to aid implementation and ensure resilience. Data warehouse developers and architects will find this book a tremendous resource for moving their skills into the future through cloud-based implementations. What You Will Learn Choose the appropriate Azure SQL engine for implementing a given data warehouse Develop smart, reusable ETL/ELT processes that are resilient and easily maintained Automate mundane development tasks through tools such as PowerShell Ensure consistency of data by creating and enforcing data contracts Explore streaming and event-driven architectures for data ingestion Create advanced staging layers using

## Get Free The Modern Data Warehouse A New Approach For A New Era

Azure Data Lake Gen 2 to feed your data warehouse Who This Book Is For Data warehouse or ETL/ELT developers who wish to implement a data warehouse project in the Azure cloud, and developers currently working in on-premise environments who want to move to the cloud, and for developers with Azure experience looking to tighten up their implementation and consolidate their knowledge

Discover how to build a cloud-based data warehouse at petabyte-scale that is burstable and built to scale for end-to-end analytical solutions Key Features Discover how to translate familiar data warehousing concepts into Redshift implementation Use impressive Redshift features to optimize development, productionizing, and operations processes Find out how to use advanced features such as concurrency scaling, Redshift Spectrum, and federated queries Book Description Amazon Redshift is a fully managed, petabyte-scale AWS cloud data warehousing service. It enables you to build new data warehouse workloads on AWS and migrate on-premises traditional data warehousing platforms to Redshift. This book on Amazon Redshift starts by focusing on Redshift architecture, showing you how to perform database administration tasks on Redshift. You'll then learn how to optimize your data warehouse to quickly execute complex analytic queries against very large datasets. Because of the massive amount of data involved in data warehousing, designing your database for analytical processing lets you take full advantage of Redshift's columnar architecture and managed services. As you advance, you'll discover how to deploy fully automated and highly scalable extract, transform, and load (ETL) processes, which help minimize the operational efforts that you have to invest in managing regular ETL pipelines and ensure the timely and accurate refreshing of your data warehouse. Finally, you'll gain a clear understanding of Redshift use cases, data ingestion, data management, security, and scaling so that you can build a scalable data warehouse platform. By the end of this Redshift book, you'll be able to implement a Redshift-based data analytics solution and have understood the best practice solutions to commonly faced problems. What you will learn Use Amazon Redshift to build petabyte-scale data warehouses that are agile at scale Integrate your data warehousing solution with a data lake using purpose-built features and services on AWS Build end-to-end analytical solutions from data sourcing to consumption with the help of useful recipes Leverage Redshift's comprehensive security capabilities to meet the most demanding business requirements Focus on architectural insights and rationale when using analytical recipes Discover best practices for working with big data to operate a fully managed solution Who this book is for This book is for anyone involved in architecting, implementing, and optimizing an Amazon Redshift data warehouse, such as data warehouse developers, data analysts, database administrators, data engineers, and data scientists. Basic knowledge of data warehousing, database systems, and cloud concepts and familiarity with Redshift will be beneficial. The new edition of the classic bestseller that launched the data warehousing industry covers new approaches and

## Get Free The Modern Data Warehouse A New Approach For A New Era

technologies, many of which have been pioneered by Inmon himself In addition to explaining the fundamentals of data warehouse systems, the book covers new topics such as methods for handling unstructured data in a data warehouse and storing data across multiple storage media Discusses the pros and cons of relational versus multidimensional design and how to measure return on investment in planning data warehouse projects Covers advanced topics, including data monitoring and testing Although the book includes an extra 100 pages worth of valuable content, the price has actually been reduced from \$65 to \$55

The data lake is a daring new approach for harnessing the power of big data technology and providing convenient self-service capabilities. But is it right for your company? This book is based on discussions with practitioners and executives from more than a hundred organizations, ranging from data-driven companies such as Google, LinkedIn, and Facebook, to governments and traditional corporate enterprises. You'll learn what a data lake is, why enterprises need one, and how to build one successfully with the best practices in this book. Alex Gorelik, CTO and founder of Waterline Data, explains why old systems and processes can no longer support data needs in the enterprise. Then, in a collection of essays about data lake implementation, you'll examine data lake initiatives, analytic projects, experiences, and best practices from data experts working in various industries. Get a succinct introduction to data warehousing, big data, and data science Learn various paths enterprises take to build a data lake Explore how to build a self-service model and best practices for providing analysts access to the data Use different methods for architecting your data lake Discover ways to implement a data lake from experts in different industries

In *Designing Cloud Data Platforms*, Danil Zburivsky and Lynda Partner reveal a six-layer approach that increases flexibility and reduces costs. Discover patterns for ingesting data from a variety of sources, then learn to harness pre-built services provided by cloud vendors. Summary Centralized data warehouses, the long-time defacto standard for housing data for analytics, are rapidly giving way to multi-faceted cloud data platforms. Companies that embrace modern cloud data platforms benefit from an integrated view of their business using all of their data and can take advantage of advanced analytic practices to drive predictions and as yet unimagined data services. *Designing Cloud Data Platforms* is a hands-on guide to envisioning and designing a modern scalable data platform that takes full advantage of the flexibility of the cloud. As you read, you'll learn the core components of a cloud data platform design, along with the role of key technologies like Spark and Kafka Streams. You'll also explore setting up processes to manage cloud-based data, keep it secure, and using advanced analytic and BI tools to analyze it. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Well-designed pipelines, storage systems, and APIs eliminate the complicated scaling and maintenance required with on-prem data centers. Once you

## Get Free The Modern Data Warehouse A New Approach For A New Era

learn the patterns for designing cloud data platforms, you'll maximize performance no matter which cloud vendor you use. About the book In Designing Cloud Data Platforms, Danil Zburivsky and Lynda Partner reveal a six-layer approach that increases flexibility and reduces costs. Discover patterns for ingesting data from a variety of sources, then learn to harness pre-built services provided by cloud vendors. What's inside Best practices for structured and unstructured data sets Cloud-ready machine learning tools Metadata and real-time analytics Defensive architecture, access, and security About the reader For data professionals familiar with the basics of cloud computing, and Hadoop or Spark. About the author Danil Zburivsky has over 10 years of experience designing and supporting large-scale data infrastructure for enterprises across the globe. Lynda Partner is the VP of Analytics-as-a-Service at Pythian, and has been on the business side of data for over 20 years. Table of Contents 1 Introducing the data platform 2 Why a data platform and not just a data warehouse 3 Getting bigger and leveraging the Big 3: Amazon, Microsoft Azure, and Google 4 Getting data into the platform 5 Organizing and processing data 6 Real-time data processing and analytics 7 Metadata layer architecture 8 Schema management 9 Data access and security 10 Fueling business value with data platforms

For undergraduate/graduate-level Data Mining or Data Warehousing courses in Information Systems or Operations Management Departments electives. Taking a multidisciplinary user/manager approach, this text looks at data warehousing technologies necessary to support the business processes of the twenty-first century. Using a balanced professional and conversational approach, it explores the basic concepts of data mining, warehousing, and visualization with an emphasis on both technical and managerial issues and the implication of these modern emerging technologies on those issues. Data mining and visualization exercises using an included fully-enabled, but time-limited version of Megaputer's PolyAnalyst and TextAnalyst data mining and visualization software give students hands-on experience with real-world applications.

Cloud Analytics with Microsoft Azure enables you to understand the design and business considerations that you must keep in mind while planning to adopt the cloud analytics model for your business.

Direct from Microsoft, this Exam Ref is the official study guide for the new Microsoft DP-200 Implementing an Azure Data Solution certification exam. Exam Ref DP-200 Implementing an Azure Data Solution offers professional-level preparation that helps candidates maximize their exam performance and sharpen their skills on the job. It focuses on specific areas of expertise modern IT professionals need to demonstrate real-world mastery of data solution provisioning, data ingestion and transformation, security, data governance, and performance monitoring. Coverage includes: Understanding Azure data solutions: data storage and data processing concepts Implementing Azure data storage solutions: implementing non-relational and relational data stores, and managing data security Managing and developing data processing for Azure data solutions: developing batch processing and streaming solutions Monitoring and optimizing Azure data solutions: monitoring data storage and data processing; troubleshooting data partitioning bottlenecks; optimizing Data Lake Storage, Stream Analytics, and Azure

## Get Free The Modern Data Warehouse A New Approach For A New Era

Synapse Analytics; and managing the data lifecycle Microsoft Exam Ref publications stand apart from third-party study guides because they: Provide guidance from Microsoft, the creator of Microsoft certification exams Target IT professional-level exam candidates with content focused on their needs, not "one-size-fits-all" content Streamline study by organizing material according to the exam's objective domain (OD), covering one functional group and its objectives in each chapter Offer concise overviews of every skill covered by the exam Feature "Thought Experiments" and "Thought Experiment Answers" to guide candidates through a set of "what if?" scenarios, and prepare them more effectively for Pro-level style exam questions Explore big picture thinking around the planning and design aspects of the IT pro's job role Deliver exam tips, summaries, and inline questions and answers to help you identify key points Include "Need more review?" reader aids pointing to additional study materials when readers need them For more information on Exam DP-200 and the Microsoft Certified: Azure Data Engineer credential, visit <https://docs.microsoft.com/en-us/learn/certifications/azure-data-engineer>

DW 2.0: The Architecture for the Next Generation of Data Warehousing is the first book on the new generation of data warehouse architecture, DW 2.0, by the father of the data warehouse. The book describes the future of data warehousing that is technologically possible today, at both an architectural level and technology level. The perspective of the book is from the top down: looking at the overall architecture and then delving into the issues underlying the components. This allows people who are building or using a data warehouse to see what lies ahead and determine what new technology to buy, how to plan extensions to the data warehouse, what can be salvaged from the current system, and how to justify the expense at the most practical level. This book gives experienced data warehouse professionals everything they need in order to implement the new generation DW 2.0. It is designed for professionals in the IT organization, including data architects, DBAs, systems design and development professionals, as well as data warehouse and knowledge management professionals. \* First book on the new generation of data warehouse architecture, DW 2.0. \* Written by the "father of the data warehouse", Bill Inmon, a columnist and newsletter editor of The Bill Inmon Channel on the Business Intelligence Network. \* Long overdue comprehensive coverage of the implementation of technology and tools that enable the new generation of the DW: metadata, temporal data, ETL, unstructured data, and data quality control.

Master the most agile and resilient design for building analytics applications: the Unified Star Schema (USS) approach. The USS has many benefits over traditional dimensional modeling. Witness the power of the USS as a single star schema that serves as a foundation for all present and future business requirements of your organization. Data warehouse legend Bill Inmon and business intelligence innovator, Francesco Puppini, explain step-by-step why the Unified Star Schema is the recommended approach for business intelligence designs today, and show through many examples how to build and use this new solution. This book contains two parts. Part I, Architecture, explains the benefits of data marts and data warehouses, covering how organizations progressed to their current state of analytics, and to the challenges that result from current business intelligence architectures. Chapter 1 covers the drivers behind and the characteristics of the data warehouse and data mart. Chapter 2 introduces dimensional modeling concepts, including fact tables, dimensions, star joins, and snowflakes. Chapter 3 recalls the evolution of the data mart. Chapter 4 explains Extract, Transform, and Load (ETL), and the value ETL brings to reporting. Chapter 5 explores the Integrated Data Mart Approach, and Chapter 6 explains how to monitor this environment. Chapter 7 describes the different types of metadata within the data warehouse environment. Chapter 8 progresses through the evolution to our current modern data warehouse environment. Part II, the Unified Star Schema, covers the Unified Star Schema (USS) approach and how it solves the challenges introduced in Part I. There are eight chapters within Part II: - Chapter 9. Introduction to the Unified Star Schema: Learn about its architecture

## Get Free The Modern Data Warehouse A New Approach For A New Era

and use cases, as well as how the USS approach differs from the traditional approach. · Chapter 10, Loss of Data: Learn about the loss of data and the USS Bridge. Understand that the USS approach does not create any join, and for this reason, it has no loss of data. · Chapter 11, The Fan Trap: Get introduced to the Oriented Data Model convention, and learn the dangers of a fan trap through an example. Differentiate join and association, and realize that an “in-memory association” is the preferred solution to the fan trap. · Chapter 12, The Chasm Trap: Become familiar with the Cartesian product, and then follow along with an example based on LinkedIn, which illustrates that a chasm trap produces unwanted duplicates. See that the USS Bridge is based on a union, which does not create any duplicates. · Chapter 13, Multi-Fact Queries: Distinguish between multiple facts “with direct connection” versus multiple facts “with no direct connection”. Explore how BI tools are capable of building aggregated virtual rows. · Chapter 14, Loops: Learn more about loops and five traditional techniques to solve them. Follow along with an implementation, which will illustrate the solution based on the USS approach. · Chapter 15, Non-Conformed Granularities: Learn about non-conformed granularities, and learn that the Unified Star Schema introduces a solution called “re-normalization”. · Chapter 16, Northwind Case Study. Witness how easy it is to detect the pitfalls of Northwind using the ODM convention. Follow along with an implementation of the USS approach on the Northwind database with various BI tools.

Updated new edition of Ralph Kimball's groundbreaking book on dimensional modeling for data warehousing and business intelligence! The first edition of Ralph Kimball's *The Data Warehouse Toolkit* introduced the industry to dimensional modeling, and now his books are considered the most authoritative guides in this space. This new third edition is a complete library of updated dimensional modeling techniques, the most comprehensive collection ever. It covers new and enhanced star schema dimensional modeling patterns, adds two new chapters on ETL techniques, includes new and expanded business matrices for 12 case studies, and more. Authored by Ralph Kimball and Margy Ross, known worldwide as educators, consultants, and influential thought leaders in data warehousing and business intelligence. Begins with fundamental design recommendations and progresses through increasingly complex scenarios. Presents unique modeling techniques for business applications such as inventory management, procurement, invoicing, accounting, customer relationship management, big data analytics, and more. Draws real-world case studies from a variety of industries, including retail sales, financial services, telecommunications, education, health care, insurance, e-commerce, and more. Design dimensional databases that are easy to understand and provide fast query response with *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition*.

A practical guide to implementing a scalable and fast state-of-the-art analytical data estate. Key Features: Store and analyze data with enterprise-grade security and auditing. Perform batch, streaming, and interactive analytics to optimize your big data solutions with ease. Develop and run parallel data processing programs using real-world enterprise scenarios. Book Description: Azure Data Lake, the modern data warehouse architecture, and related data services on Azure enable organizations to build their own customized analytical platform to fit any analytical requirements in terms of volume, speed, and quality. This book is your guide to learning all the features and capabilities of Azure data services for storing, processing, and analyzing data (structured, unstructured, and semi-structured) of any size. You will explore key techniques for ingesting and storing data and performing batch, streaming, and interactive analytics. The book also shows you how to overcome various challenges and complexities relating to productivity and scaling. Next, you will be able to develop and run massive data workloads to perform different actions. Using a cloud-based big data-modern data warehouse-analytics setup, you will also be able to build secure, scalable data estates for enterprises. Finally, you will not only learn how to develop a data warehouse but also understand how to create enterprise-grade security and auditing big data programs. By the end of this Azure book, you will have learned how to develop a

## Get Free The Modern Data Warehouse A New Approach For A New Era

powerful and efficient analytical platform to meet enterprise needs. What you will learn Implement data governance with Azure services Use integrated monitoring in the Azure Portal and integrate Azure Data Lake Storage into the Azure Monitor Explore the serverless feature for ad-hoc data discovery, logical data warehousing, and data wrangling Implement networking with Synapse Analytics and Spark pools Create and run Spark jobs with Databricks clusters Implement streaming using Azure Functions, a serverless runtime environment on Azure Explore the predefined ML services in Azure and use them in your app Who this book is for This book is for data architects, ETL developers, or anyone who wants to get well-versed with Azure data services to implement an analytical data estate for their enterprise. The book will also appeal to data scientists and data analysts who want to explore all the capabilities of Azure data services, which can be used to store, process, and analyze any kind of data. A beginner-level understanding of data analysis and streaming will be required.

[Copyright: 8f893abb4de97249874ac0c3f453e652](#)